Responsible Disclosure of Generative Models Using Scalable Fingerprinting

Year: 2020 | Citations: 105 | Authors: Ning Yu, Vladislav Skripniuk, Dingfan Chen, Larry S. Davis, Mario Fritz

---

Abstract

Over the past five years, deep generative models have achieved a qualitative new level of performance. Generated data has become difficult, if not impossible, to be distinguished from real data. While there are plenty of use cases that benefit from this technology, there are also strong concerns on how this new technology can be misused to spoof sensors, generate deep fakes, and enable misinformation at scale. Unfortunately, current deep fake detection methods are not sustainable, as the gap between real and fake continues to close. In contrast, our work enables a responsible disclosure of such state-of-the-art generative models, that allows researchers and companies to fingerprint their models, so that the generated samples containing a fingerprint can be accurately detected and attributed to a source. Our technique achieves this by an efficient and scalable ad-hoc generation of a large population of models with distinct fingerprints. Our recommended operation point uses a 128-bit fingerprint which in principle results in more than 10^{36} identifiable models. Experimental results show that our method fulfills key properties of a fingerprinting mechanism and achieves effectiveness in deep fake detection and attribution.